

CLAIMS:

What is claimed is:

1. A method of computing a model, comprising:
comparing attribute values for samples having a desired attribute to attribute values for all samples; and
selecting a subset of available attributes based on a difference between attribute values for the samples having the desired attribute and attribute values for all of the samples.

2. The method of claim 1, wherein the step of comparing attribute values for samples having a desired attribute to attribute values for all samples further comprises:
determining a statistical measure of difference between the attribute values for samples having the desired attribute and the attribute values for all of the samples.

3. The method of claim 2, wherein the step of determining a statistical measure of difference between the attribute values for samples having the desired attribute and the attribute values for all of the samples further comprises:
determining an entropy for the attribute values.

4. The method of claim 1, wherein the step of selecting a subset of available attributes based on a difference between attribute values for the samples having the desired attribute and attribute values for all of the samples further comprises:
identifying n attributes having a largest difference in attribute values.

1 5. The method of claim 1, wherein the step of selecting a
2 subset of available attributes based on a difference between
3 attribute values for the samples having the desired
4 attribute and attribute values for all of the samples
5 further comprises:
6

7 identifying a predetermined percentage of attributes
8 having a larger difference in the attribute values than
remaining attributes.

1 6. The method of claim 1, wherein the step of selecting a
2 subset of available attributes based on a difference between
3 attribute values for the samples having the desired
4 attribute and attribute values for all of the samples
5 further comprises:
6

7 identifying attributes having a difference in the
8 attribute values exceeding a predetermined amount.

1 7. The method of claim 1, further comprising:
2 obtaining a plurality of samples, each sample having
3 values for a plurality of attributes.
4

5 8. The method of claim 1, further comprising:
6 employing the selected subset of attributes to generate
7 a predictive model.
8

1 9. A method of computing a model, comprising:
2 obtaining a plurality of samples each having values for
3 a plurality of attributes;
4 comparing attribute values for samples having at least
5 one desired attribute to attribute values for all of the
6 plurality of samples;
7 selecting attributes having a largest difference
8 between attribute values for samples having the at least one
9 desired attribute and attribute values for all of the
10 plurality of samples; and
11 computing a model employing the selected attributes.

10 10. The method of claim 9, wherein the step of selecting
11 attributes having a largest difference between attribute
12 values for samples having the at least one desired attribute
13 and attribute values for all of the plurality of samples
14 further comprises:

15 identifying a predetermined number of attributes having
16 the largest difference in attribute values.

17 11. The method of claim 9, wherein the step of selecting
18 attributes having a largest difference between attribute
19 values for samples having the at least one desired attribute
20 and attribute values for all of the plurality of samples
21 further comprises:

22 identifying a predetermined percentage of attributes
23 having the relative difference in attribute values.

24 12. The method of claim 9, wherein the step of selecting
25 attributes having a largest difference between attribute
26 values for samples having the at least one desired attribute
27 and attribute values for all of the plurality of samples

5 further comprises:
6 identifying attributes having a difference in attribute
7 values equal to or greater than a predetermined amount.

09282619 033199
66665.1

1 13. A method of selecting attributes for computing a model,
2 comprising:

3 for a plurality of samples each having values for a
4 plurality of attributes:

5 for each of the plurality of attributes:

6 comparing the attribute values for a first
7 group of samples to the attribute values for all
8 of the plurality of samples; and

9 determining a difference between the
10 attribute values for the first groups and the
11 attribute values for all of the plurality of
12 samples; and

13 identifying attributes within the plurality of
14 attributes having a largest difference between the
15 attribute values for the first groups and the attribute
16 values for all of the plurality of samples; and
17 selecting at least some of the identified attributes.

1 14. A system for selecting attributes for computing a
2 model, comprising:
3 a memory containing data for a plurality of samples
4 each having values for a plurality of attributes; and
5 a processor coupled to the memory and executing a
6 selection process including:
7 comparing attribute values for samples having a
8 desired attribute to attribute values for all samples;
9 selecting a subset of available attributes based
10 on a difference between attribute values for the
11 samples having the desired attribute and attribute
12 values for all of the samples; and
13 employing the selected subset of attributes to generate
14 a predictive model.

15. The system of claim 14, wherein the selection process
determines a statistical measure of difference between the
attribute values for samples having the desired attribute
and the attribute values for all of the samples.

16. The system of claim 15, wherein the selection process
determines an entropy for the attribute values.

17. The system of claim 14, wherein the selection process
identifies a predetermined number of attributes having a
largest difference in the attribute values for selection.

18. The system of claim 14, wherein the selection process
identifies a predetermined percentage of attributes having a
larger difference in the attribute values for selection.

19. The system of claim 14, wherein the selection process

- 2 identifies, for selection, attributes having a difference in
3 ~~the attribute values exceeding a predetermined amount.~~

48

66665.1

1 20. A system for computing a model, comprising:
2 a memory containing data for a plurality of samples
3 each having values for a plurality of attributes; and
4 a processor coupled to the memory and executing a
5 selection process including:
6 comparing attribute values for a first subset of
7 the plurality of samples to attribute values for all of
8 the samples;
9 selecting attributes having a largest difference
10 between attribute values for the first subset and
11 attribute values for all of the samples; and
12 computing a model employing the selected
13 attributes.

1 21. A computer program product within a computer usable
2 medium for selecting attributes for computing a model,
3 comprising:

4 instructions for reading values of attributes for a
5 plurality of samples;

6 instructions for comparing attribute values for samples
7 having a desired attribute to attribute values for all
8 samples; and

9 instructions for selecting a subset of available
10 attributes based on a difference between attribute values
11 for samples having the desired attribute and attribute
12 values for all samples.

13 22. The computer program product of claim 21, wherein the
14 instructions for comparing attribute values for samples
15 having a desired attribute to attribute values for all
16 samples further comprise:

17 instructions for determining a statistical measure of
18 difference between the attribute values for samples having
19 the desired attribute and the attribute values for all
20 samples.

21 23. The computer program product of claim 22, wherein the
22 instructions for determining a statistical measure of
23 difference between the attribute values for samples having
24 the desired attribute and the attribute values for all
25 samples further comprise:

26 instructions for determining an entropy of the
27 attribute values for samples having the desired attribute
28 and an entropy of the attribute values for all samples;

29 instructions for comparing the entropy of the attribute
30 values for samples having the desired attribute to the

11 entropy of the attribute values for all samples for each
12 attribute to determine a relative measure of difference; and
13 instructions for comparing the relative measure of
14 difference of all attributes.
15

1 24. The computer program product of claim 21, wherein the
2 instructions for selecting a subset of available attributes
3 based on a difference between attribute values for samples
4 having the desired attribute and attribute values for all
5 samples further comprise:

6 instructions for identifying n attributes having a
7 largest difference in the attribute values.

1 25. A computer program product within a computer usable
2 medium for selecting attributes for computing a model,
3 comprising:
4

5 instructions for comparing attribute values for a first
6 group of samples to attribute values for all samples for
7 each of a plurality of attributes;

8 instructions for determining a difference between the
9 attribute values for the first group of samples and the
10 attribute values for all of the samples; and

11 instructions for selecting a group of attributes having
12 a largest difference between the attribute values for the
13 first group of samples and the attribute values for all
14 samples.